(19) Europäisches Pat ntamt
European Patent Offic
Office européen d s brevets

(11) EP 0 987 894 A2

(12) EUROPEAN PATENT APPLICATION

(43) Date of publication:
22.03.2000 Bulletin 2000/12

(51) Int Cl.[7]: H04N 7/24

(21) Application number: 99304426.2

(22) Date of filing: 08.06.1999

(84) Designated Contracting States:
AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE
Designated Extension States:
AL LT LV MK RO SI

(30) Priority: 07.09.1998 SG 9803578

(71) Applicant: VICTOR COMPANY OF JAPAN, LTD.
Yokohama-Shi, Kanagawa-Ken (JP)

(72) Inventor: XU, Jinghong
Singapore 760247 (SG)

(74) Representative: Crawford, Andrew Birkby et al
A.A. Thornton & Co.
235 High Holborn
London WC1V 7LE (GB)

(54) A dejittering and clock recovery technique for real-time audio/visual network applications

(57) In a real-time audio/visual system in which A/V data is conveyed over a jitter-introducing network, dejittering and clock recovery processes can be achieved without requiring a Phase Locked Loop (PLL). At the server, audio/video streams are encoded into transport packets before being sent out. At the client, the dejittering process is achieved by a dejittering buffer using the embedded timestamps in the transport packets and a client decoding clock. The delay variations of data arriving are removed after the client buffering process. At the scheduled time, each data packet is shifted to a synchronizing buffer and then fed to the A/V decoder according to the speed of A/V stream. The clock synchronization between client and server is achieved by a synchronizing buffer whose half-size position is taken as the reference. By monitoring the movement of the buffer fill position over a given period, the drift rate of clock unsynchronization between client and server can be derived and, therefore, the client's clock can be adjusted to synchronize with the server's clock based on the derived drift.

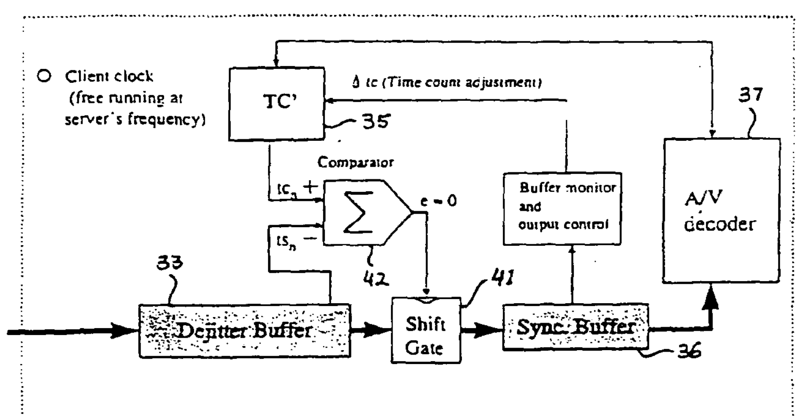

Figure 2

EP 0 987 894 A2

Printed by Jouve, 75001 PARIS (FR)

Description

Field of the Invention

[0001] This invention relates to a method for network dejittering and client/server timing synchronization in network (such as IP network) applications for real-time audio/visual services. This technique may be used for the A/V stream (i.e. MPEG stream) decoder to smooth the network jitter and to slave its timing to the encoder by a software approach. With this method, a client can playback the A/V streams over an IP network with no requirement of a PLL (Phase Lock Loop) circuit in the decoder.

Background of the Invention

[0002] With the continuing growing of Internet and IP-related technologies, there is a growing demand for access to multimedia application across a IP-based network. Unlike the traditional point-to-point network applications, emerging multimedia applications such as live or storage distance learning, TV broadcast directly to desktop via IP network, and desktop conferencing depend on the ability of multicasting for real-time services. The rising need for these kind of applications presents the challenge for network and end-systems. It is well known that the Internet has been used primarily for the reliable transmission of conventional data with minimal or no constraints of delay. Its protocols such as TCP/IP were well designed for this type of traffic using "Pull" mode. However, multimedia data such as audio and video are delay sensitive. Such traffic possesses different characteristics and may require different protocols in order to effectively provide real-time services. For digital A/V broadcast services, it requires appropriate "Push" mode operation with large uni-directional channel, with a small (or no) return channel required.

[0003] As one of the main requirements for an IP networked digital audio/visual system, people expect that the system will allow the ability to receive or playback audio/visual information over the IP network in a real-time fashion. However, due to the nature of the IP, low layer network access and platform dependant system clock, three main problem issues must be faced. One problem is the quality of service (QoS). A second problem is that there is a significant amount of delay variation (jitter) presented in an end-to-end system over IP network. When an MPEG system stream is transmitted over a jitter-inducing network, the real byte data delivery schedule may differ significantly from the intended delivery schedule. In such a situation, it is not possible to decode the system stream on a standard target decoder, because jitter may cause buffer overflows or underflows and also make it difficult to recover the time base. The third problem is the unmatched clocks (time drift) between server (encoder) and client (decoder). The time drift means that the free-running clock in the client will introduce some amount of timing difference over a given period of time when compared to the server's encoding clock since there is no guarantee of clock accuracy. The first problem relates to the issue of how to guarantee the bandwidth and delay for the data delivery over an IP network and is currently addressed by the IETF. Available technologies addressing this problem are: Resource Reservation Protocol (RSVP), Differential Service, Multi-Protocol Label Switching (MPLS), etc. The embodiments of the present invention are intended for addressing the other two problems. These two problems cause the difficulty for clients to decode data accurately and playback in a real-time mode using conventional technology.

[0004] The problem of using the existed technology in such a system can be illustrated by the example described below. For example, within an MPEG-2 system data stream there are clock time reference timestamps. These references are samples of the system time clock and have a resolution of one part in 27,000,000 per second. They occur at intervals up to 100 ms in Transport Streams (TS) or up to 700 ms in Program Streams (PS). In the PS, the clock field is called the System Clock Reference (SCR). In the TS, it is called the Program Clock Reference (PCR). The SCR (or PCR) field indicates the correct value of the STC (System Time Clock - a common time base) when the SCR (or PCR) is received at the decoder. In concept. this STC value is the same value that the encoder's STC had when the SCR (or PCR) was stored or transmitted. If the decoder's clock frequency matches exactly that of the encoder, then the decoding and presentation of audio/video will automatically have the same rate as those at the encoder. In practice, a decoder's system clock frequency will not precisely match the encoder's system clock frequency.

[0005] The decoder's STC can be made to slave its timing to the encoder using the received SCRs (or PCRs). The prototypical method of achieving such synchronization is via a Phase-Locked Loop (PLL). The details on the operation of PLL in this context may be found in ISO/IEC International Standard 13818-1. "Generic Coding of Moving Pictures and Associated Audio Information: Systems", July 1995. There is a bounded maximum interval between successive SCRs (or PCRs) in order to allow the operation of PLL to be stable. In a jitter introducing network, packet delay variation is considered to be quite significant for the MPEG Standard Target Decoder (STD). Such timing jitter at the input to a decoder is reflected in the combination of the values of SCRs (or PCRs) and the times when they are received. A significant SCRs (PCRs) jitter may cause the difficulty for the PLL to reach a defined locked state. The effect of such clock frequency mismatch between client and server is the gradual and unavoidable increase or decrease of the fullness of the decoder's buffers, such that overflow or underflow would occur eventually with any finite size of decoder buffers, therefore, effecting the system performance of the overall operation.

[0006] It is an object of the present invention, to provide a technique for dejittering and clock recovery for use in the client application of a networked real-time audio/visual service system. It is desirable to be able to implement embodiments of the invention to synchronize the clients to a server in a jitter introducing network environment without employing additional devices or a special decoder.

Summary of the Invention

[0007] In accordance with the present invention, there is provided a method for clock variation compensation in a real-time audio/visual system in which encoded A/V data in a plurality of data packets are delivered to at least one client over a network from a server at a substantially constant bit rate, said plurality of data packets including data packets containing time stamp data the method comprising the steps of:

receiving and buffering the data packets in a first buffer at a client;
passing selected data packets from the first buffer to a second buffer at scheduled times based on a comparison between a local clock of the client and timestamp data corresponding to the selected data packets; and
passing the data in the second buffer to a data decoder of the client.

[0008] Preferably the method includes monitoring the fullness of the second buffer to derive a drift rate, and adjusting the client local clock based on the derived drift rate.

[0009] The present invention also provides a method for clock variation compensation in a real-time audio/visual system in which encoded A/V data in a plurality of data packets are delivered to at least one client over a network from a server at a substantially constant bit rate, said plurality of data packets including data packets containing timestamp data, the method comprising the steps of:

receiving and buffering the data packets in a dejittering buffer at a client;
passing selected data packets from the dejittering buffer to a data decoder of the client at scheduled times based on a comparison between a decoding system clock of the client and timestamp data corresponding to the selected data packets.

[0010] Preferably the method includes monitoring the fullness of the dejittering buffer to derive a clock drift rate, and adjusting the decoding system clock based on the derived drift rate and a network jitter component.

[0011] The present invention further provides a real-time audio/visual system coupled to receive data packets over a network for A/V decoding by an A/V decoder, th system including a clock variation compensation system comprising:

a dejitter buffer for receiving and storing packets of data from the network;
a synchronization buffer for feeding data for decoding to the A/V decoder;
a decoder system clock; and
a buffer data flow controller for controlling the passing of selected data packets from the dejitter buffer to the synchronization buffer in accordance with a comparison of a first signal derived from the decoder system clock and a second signal derived from a timestamp from the selected data packets.

[0012] The technique summarized below is a "software PLL-like" method to address the jitter and time synchronization for a client decoding system. It employs the RTP (Real-Time Transport Protocol) as the transport service and receiver buffering to achieve real-time A/V playback.

[0013] The dejittering process can be achieved by a dejittering buffer using the embedded timestamp values in the transport packets and client RTP clock (which runs at the same frequency as the A/V decoder's clock). The delay variations of data arriving are removed after the client buffering process, the data packet is shifted to a synchronizing buffer at the scheduled time of server encoder, then feed to the A/V decoder according to the time reference of A/V encoder. The clock synchronization (recovery) can be achieved by a synchronizing buffer based on a reference fill position and the movement of the fill position in the buffer (packet index oriented). By monitoring the fullness of the buffer over a given period, the drift rate of clock unsynchronization between client and server can be derived.

Brief Description of the Drawings

[0014] The invention is described in greater detail hereinafter, by way of example only, with reference to embodiments thereof which are described with the aid of the accompanying drawings, wherein:

Figure 1 schematically illustrates a networked system for real-time audio/visual services comprising a plurality of client host and a server;
Figure 2 schematically illustrates a client host comprising a dejittering buffer and a software PLL-like architecture, wherein dejittering and clock recovering is achieved in accordance with an embodiment of the present invention; and
Figure 3 diagrammatically illustrates the mechanism of timing synchronization via monitoring the position movement of the dejittering buffer.

Detailed Description of the Preferred Embodiments

[0015] In this specification, various aspects of the Real-Time Transport Protocol (RTP) and Resource Reservation Protocol (RSVP) are referred to, and a more detailed description of the mechanisms, protocols and systems involved in implementing RTP and RSVP can be found in the following documents, the disclosures of which are incorporated herein by reference:
"RTP: A Transport Protocol for Real-Time Applications", H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson; RFC 1889, January 1996.
"RTP Profile for Audio and Video Conferences with Minimal Control", H. Schulzrinne; RFC 1890, January 1996.
"RTP Payload Format for MPEG1/MPEG2 Video", D. Hoffman, G. Fernando, V. Goyal, M.R. Civanlar; drat-ietf-avt-mpeg-new-01, Internet draft, June 1997.
"RTP Payload Format for Bundled MPEG", M.R. Civanlar, G.L Cash, B.G Haskell; draft-civanlar-bmpeg-01, Internet draft, February 1997.
"Resource ReSerVation Protocol (RSVP) Version 1 Function Specification", R. Braden, L. Zhang, S. Berson, S. Herzog, S Jamin, draft-ietf-rsvp-spec-16, Internet draft, June 1997.

[0016] A system model for real-time services over a jitter introducing network is illustrated in block diagram form in Figure 1. In the Figure, an A/V client server arrangement is shown, wherein a client 30 is connected to a sever 10 via a network 20. The server 10 includes a source of audio/video streams 11, a RTP encoder 12, a server encoding system clock 13, transport and network layers 14 and a network interface 15. The client 30 includes a network interface 31, transport and network layers 32, a dejittering buffer 33, a RTP decoder 34, a client decoding system clock 35, a time synchronizing buffer 36 and an audio/video decoder 37. The network 20 is a jitter introducing network, such as IP-based network. The overall operation of the present invention in this environment is described hereinbelow.

[0017] At the server 10, AV stream source 11 is input to the RTP encoder 12. The output of RTP encoder 12 are RTP packets where their payload field contain data from source 11. These packets are timestamped by the server system encoding clock 13, and then sent out to the client through the network 20 at constant bit rate. At the client 30, the RTP packets (containing AV stream data in payload field) received from the network are put into the dejittering buffer 33 in sequence. The RTP packets in dejittering buffer 33 are decoded by the RTP decoder 34 and client decoding system clock 35, and then shifted into the time synchronizing buffer 36. In other words, the synchronizing buffer 36 will contain all the original A/V stream data from A/V source 11. These A/V stream data will then be fed into A/V decoder according to their appropriate time scheduling.

[0018] The desirable size of the dejittering buffer depends only on the maximum network jitter J_{max} (peak-to-peak) and the maximum bit rate of A/V streams R_{max}. Assume the difference of clock counting between the server and client over the period of updating (*T*, which for example can be default to 1 minute) is *t*. The size of the dejittering buffer can be determined by:

$$B_{dj} = J_{max} * (R_{max} + t)$$

For example, if J_{max} = 100ms, R_{max} = 8Mbps and t = 10 ms, then the required minimum buffer size is,

$$B_{dj} = 8 \times 10^6 \times (100 + 10) \times 10^3/8$$

$$= 110000 \text{ Bytes} < 125 \text{ kBytes}$$

[0019] Therefore, a size of 512 kBytes buffer should be adequate for most situations if a certain level of QoS is also guaranteed.

[0020] Figure 2 is a block diagram of the dejittering and clock recovering operation in accordance with an embodiment of the present invention. In this diagram, and the following description:

TC'	is the time counter of the client's RTP clock
tc_n	is the time counter value for the n^{th} packet
ts_n	is the timestamp value of the n^{th} packet in the dejitter buffer
tc	is the adjustment value for TC, when necessary.

[0021] As detailed in the block diagram of Figure 2, the dejittering process is done based on the encoded RTP timestamp value and client's RTP clock. At the first step, the client buffer provides the function of buffering by way of dejitter buffer 33, which removes the jitter in terms of client reception. The dejitter buffer 33 is coupled to pass data to the synchronizing buffer 36 by way of a shift gate 41. The shift gate 41 is controlled by the output of a comparator 42 which has inputs tc_n from the client's RTP clock time counter and ts_n retrieved from the dejitter buffer. The data is shifted to the synchronizing buffer 36 whenever the timestamp value (ts_n) encoded in the packet is equal to the client's time counter value (tc_n). The packets in the synchronizing buffer are A/V data packets. Such data packets are fed to the A/V decoder (e.g. MPEG decoder) based on their own embedded time schedule. It should be noted here that in the case of a system decoder the decoder internal system buffer can be used as the synchronizing buffer if it is accessible from the client.

[0022] The clock synchronization between client and server is achieved via monitoring the packet position in the synchronizing buffer. The size of the synchronizing buffer can be quite small since it only needs to handle the clock drift. The operation is illustrated in Figure 3 which is a diagram of buffer monitoring for time synchronization. Here the operation of the clock recovery process is based on the buffer position change compared

with the reference position (defined as the buffer half-size position) for a given period T (for example, every minute). If the buffer position is moving in the direction of emptiness, the counter TC' should be upwardly adjusted by adding an offset. If the buffer position is moving in the other direction (fullness), the TC' value should be downwardly adjusted. The drift rate r of clock unsynchronization between server and client can be determined by:

$$r = (p_2 - p_1)/T$$

[0023] It should be noted that the buffer position mentioned above is in terms of packet index offset rather than the byte number offset.

[0024] It is also possible to implement the present technique with the use of one buffer only (no additional synchronizing buffer available). In such case, data packets are fed into the A/V decoder directly from the dejittering buffer at the scheduled time. The above clock drift rate r will include the component of network jitter J and should be eliminated before the TC' is adjusted. The interarrival jitter J can be derived from the two sequenced RTP packets. The jitter J is defined to be the mean deviation of the difference (D) in packet spacing at the receiver compared to the sender for a pair of packets. For example, if T_{sa} is the RTP timestamp from packet a, and T_{rb} is the time of arrival in RTP timestamp units for packet b, then for these two packets, we have,

$$j = D(a,b)$$

$$= (T_{rb} - T_{ra}) - (T_{sb} - T_{sa})$$

$$= (T_{rb} - T_{sb}) - (T_{ra} - T_{sa})$$

[0025] The j is calculated continuously as each data packet is received from server, then according to the formula, we have,

$$J = J + (j - J)/16$$

[0026] This algorithm provides an optimal first-order estimator and the gain parameter 1/16 gives a good noise reduction ratio while maintaining a reasonable rate of convergence (see RTP specification and related references).

[0027] Unlike the traditional method (using a PLL circuitry) and other available technique (such as that disclosed in European patent No. EP779725, entitled "Method and Apparatus for Delivering Simultaneous Constant Bit Rate Compressed Video Streams At Arbitrary Bit Rates with Constrained Drift and Jitter", which uses two levels of synchronization named coarse-grain and fine-grain for video streams to control drift and jitter), the present method is a more independent and less constrained client-based approach. It provides the network adaptability via a simple software solution (can be implemented in hardware as well) for an end system and can be applied in a wide range of network applications. The effect of the present technique includes:

- increasing the adaptability of decoding systems (in terms of dejittering and clock synchronization);
- simplifying the decoder implementation (no PLL circuitry required);
- handling more types of audio/visual streams (not only MPEG2 system streams);
- application to different types of environments (unicast as well as multicast);
- beneficial to the effort of expanding real-time A/V services over IP-based network such as Internet.

[0028] The foregoing detailed description of embodiments and implementations of the invention have been presented by way of example only, and is not intended to be considered limiting to the present invention as defined in the claims appended hereto. Numerous alternative embodiments may be devised by those skilled in the art without departing from the spirit and scope of the invention.

[0029] Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", and variations such as "comprises" and "comprising", will be understood to imply the inclusion of a stated integer or step or group of integers or steps but not the exclusion of any other integer or step or group of integers or steps.

Claims

1. A method for clock variation compensation in a real-time audio/visual system in which encoded A/V data in a plurality of data packets are delivered to at least one client over a network from a server at a substantially constant bit rate, said plurality of data packets including data packets containing timestamp data, the method comprising the steps of:

   receiving and buffering the data packets in a first buffer at a client;
   passing selected data packets from the first buffer to a second buffer at scheduled times based on a comparison between a local clock of the client and timestamp data corresponding to the selected data packets; and
   passing the data in the second buffer to a data decoder of the client.

2. A method as claimed in claim 1, including monitoring the fullness of the second buffer to derive a drift rate, and adjusting the client local clock based on

the derived drift rate.

3. A method as claimed in claim 1 or 2, wherein the scheduled time is determined by a substantially zero difference between the value of a said timestamp data corresponding to the selected data packets and a clock counter from the client local clock.

4. A method as claimed in claim 2, wherein the drift rate is derived according to:

$$r = (p2 - p1)/T$$

where:

r is the derived drift rate, and
p2 and *p1* are two corresponding buffer positions for a given period of time *T*.

5. A method for clock variation compensation in a real-time audio/visual system in which encoded A/V data in a plurality of data packets are delivered to at least one client over a network from a server at a substantially constant bit rate, said plurality of data packets including data packets containing timestamp data, the method comprising the steps of:

receiving and buffering the data packets in a dejittering buffer at a client;
passing selected data packets from the dejittering buffer to a data decoder of the client at scheduled times based on a comparison between a decoding system clock of the client and timestamp data corresponding to the selected data packets.

6. A method as claimed in claim 5, including monitoring the fullness of the dejittering buffer to derive a clock drift rate and adjusting the decoding system clock based on the derived drift rate and a network jitter component.

7. A method as claimed in claim 6, wherein said network jitter component is based on packet interarrival jitter information.

8. A method as claimed in claim 6, wherein said network jitter component is derived in accordance with the IETF Real-time Transport Protocol (RTP).

9. A method as claimed in claim 5, wherein the data packets are transmitted from said server to said client over said network in accordance with the RTP protocol.

10. A real-time audio/visual system coupled to receive data packets over a network for A/V decoding by an A/V decoder, the system including a clock variation compensation system comprising:

a dejitter buffer for receiving and storing packets of data from the network;
a synchronization buffer for feeding data for decoding to the A/V decoder;
a decoder system clock; and
a buffer data flow controller for controlling the passing of selected data packets from the dejitter buffer to the synchronization buffer in accordance with a comparison of a first signal derived from the decoder system clock and a second signal derived from a timestamp from the selected data packets.

11. A system as claimed in claim 10, further including a clock adjustment controller coupled to monitor a measure of the fullness of one of the synchronization and dejitter buffers, and coupled to the decoder system clock to allow adjustment thereof based on the buffer fullness measurement.

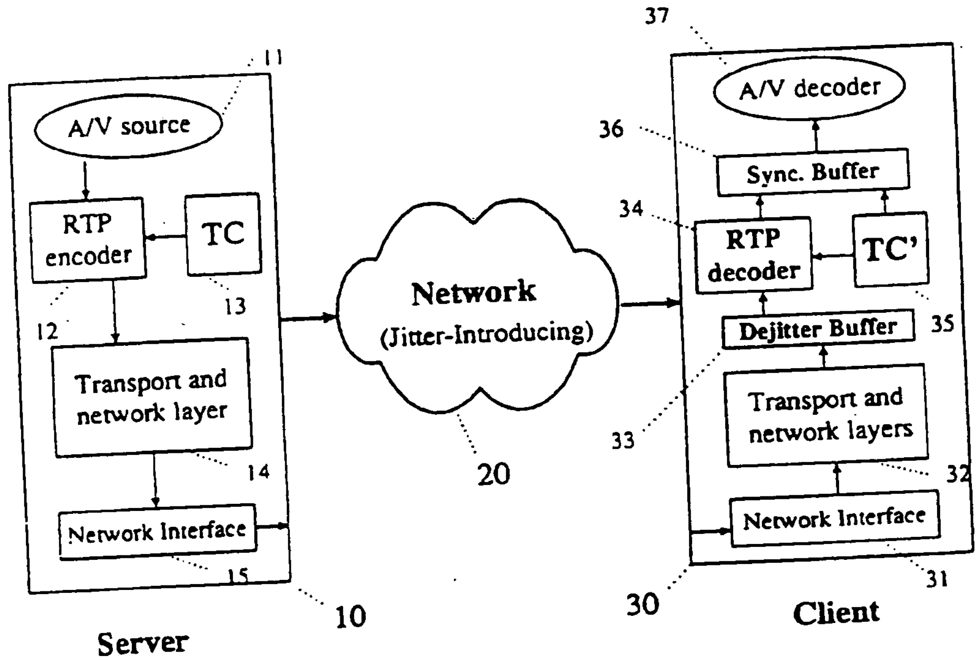
A/V source
RTP encoder
TC
Transport and network layer
Network Interface
11
12
13
14
15
10
Server
Network
(Jitter-Introducing)
20
A/V decoder
Sync. Buffer
RTP decoder
TC'
Dejitter Buffer
Transport and network layers
Network Interface
37
36
34
35
33
32
31
30
Client

Figure 1

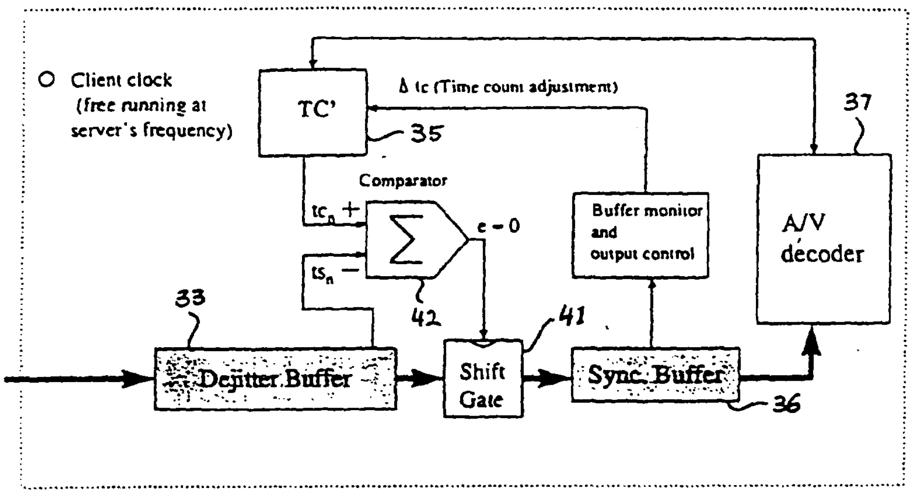

Figure 2

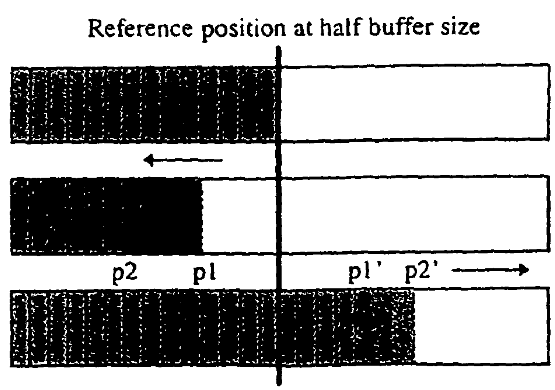

Figure 3

A dejittering and clock recovery technique for real-time audio/visual network applications

Patent Number:	EP0987894, A3
Publication dat :	2000-03-22
Inventor(s):	XU JINGHONG (SG)
Applicant(s):	VICTOR COMPANY OF JAPAN (JP)
Requested Patent:	JP2000092130
Application Number:	EP19990304426 19990608
Priority Number(s):	SG19980003578 19980907
IPC Classification:	H04N7/24
EC Classification:	H04N7/24
Equivalents:	SG71835
Cited Documents:	US5633871; WO9522233; US5790543; EP0712250; US5640388

Abstract

In a real-time audio/visual system in which A/V data is conveyed over a jitter -introducing network, dejittering and clock recovery processes can be achieved without requiring a Phase Locked Loop (PLL). At the server, audio/video streams are encoded into transport packets before being sent out. At the client, the dejittering process is achieved by a dejittering buffer using the embedded timestamps in the transport packets and a client decoding clock. The delay variations of data arriving are removed after the client buffering process. At the scheduled time, each data packet is shifted to a synchronizing buffer and then fed to the A/V decoder according to the speed of A/V stream. The clock synchronization between client and server is achieved by a synchronizing buffer whose half-size position is taken as the reference. By monitoring the movement of the buffer fill position over a given period, the drift rate of clock unsynchronization between client and server can be derived and, therefore, the client's clock can be adjusted to synchronize with the server's clock based on the derived drift.